Guanze Lu

wenqq3@outlook.com | github.com/wenqingqian | wenqingqian.github.io

Education

Hangzhou Dianzi University Graduation Time: 2024/06 (Expected)

Bachelor of Engineering (B.Eng) Hangzhou, China

- Major: Intelligent science and Technology
- GPA:

University of Chinese Academy of Sciences Admission Time: 2024/09 (Expected)

Master of Engineering

- Major: Computer Technology

Research Interests

I am broadly interested in the intersection between AI and efficient machine learning system (Mainly based on GPU). Examples include building **efficient computation systems for NLP training & inference and supercomputer scientific applications.**

Work in Progress

Graduation Design

improved CycleGAN in PaperCut

CUDA Kernel Optim

- SGEMM, SGEMV, Reduce, dot-product, online-softmax, ...
- SGEMM Reach 70% of peak performance (cuBlas 80%)

Software Projects

CPU Emulator

NEMU ICS2020 OpenCourse

- Sequential execution RISCV CPU emulator
- Include run program in OS and run OS in emulator
- Support multiprogramming, virtual memory, IO, simple filesystem, clock interrupts and exception handling mechanisms

OS Kernel

XV6 MIT 6.S081 version:2021 OpenCourse

- Time-sharing operating system
- Features (u)syscall, copy on write, large file, mmap, ...

DeepLearning System

Needle CMU 10-714 version:2022 OpenCourse

- Automatic differentiation, basic model, dataloader, CPU(C++)&GPU(CUDA) backend

Network Lib

Muduo Lib

- epoll at ET mode, asynchronous log and toml configure parse with C++20 (variant, optional, concept, ...)

Technical Skills

- **Languages:** C/C++, Python, Lua, CUDA, Shell
  - Familiar with C++11
  - Understand and have used template metaprogramming and some new features
- **Frameworks and Tools:** PyTorch, RISC-V, CMake

Last Updated on May 7, 2024